ISSN 0317-2473

Research Discussion Paper No. 73

QUANTIFYING SEMI-STRUCTURED INTERVIEWS

By

Harvey Krahn

Department of Sociology
University of Alberta
Edmonton, Alberta

Julian Tanner's useful comments on an earlier draft and Dave Odynak's assistance in constructing data files are acknowledged. Funding for this study was provided by Alberta Manpower, the City of Edmonton, the University of Alberta, and the Solicitor General of Canada.

DECEMBER 1990

92-04426/

QUANTIFYING SEMI-STRUCTURED INTERVIEWS

ABSTRACT

This paper describes an attempt to quantify textual data obtained from semi-structured interviews. A content analysis of transcripts from interviews with 162 high school dropouts was used to construct coding categories. These categories were then used to create a numerical data set which was merged with data from standardized questionnaires completed by the same subjects. Assessment of *information adequacy*, and of the *efficiency* and *research utility* of this approach, suggests that it has considerable value. However, researchers must remember that the numeric data are meant to supplement, not replace, the textual data. In addition, researchers can improve the quality of the numeric data extracted from semi-structured interviews by following several general coding guidelines.

INTRODUCTION

Most qualitative researchers would allow that when choosing a representative group of informants, or while counting the frequency of certain types of behaviours or events, they are incorporating elements of survey research. Similarly, most survey researchers would recognize the interpretive components of their quantitative studies. But sociologists have been reluctant to argue for a deliberate combination of the two methods. Much of this resistance stems from the long-standing and unproductive debate over the relative merits of "deep, rich" qualitative data and "hard, generalizeable" data from surveys (Sieber, 1973:1335). Nevertheless, in the past decades, multi-method research programs have tackled a growing range of research problems (Bryman, 1988:127-56).

The different insights and the chance to validate findings which triangulation of methods allows are widely acknowledged. However, in most multi-method studies the two approaches are still kept largely separate, and one approach typically dominates (Bryman, 1988:128). Survey researchers might interview a smaller number of key informants (e.g. Connidis, 1983). Alternatively, field researchers might undertake surveys to supplement their more extensive qualitative data collection efforts (e.g. Wallace, 1987). More contentious are efforts to actually merge methods. This paper describes and evaluates one such effort in which textual data from semi-structured interviews were quantified and linked with standardized questionnaire data.

The growing use of computers in qualitative research makes such an approach much more feasible than it would have been a decade ago. Although mainframe computers have not been used extensively by field researchers, the potential of micro-computers with their word-processing and file management capacities was quickly recognized. Micro-computers are extremely useful for marking and retrieving key passages in textual data. Electronic scanning for specific types of data, or for previously imbedded thematic markers or cross-references, is much more efficient than traditional searches through even the most carefully organized field

notes or interview transcripts. Once interview records have been classified and labelled, computers can sort, count, and cross-classify them effortlessly.

By 1984, a special issue of Qualitative Sociology could be devoted to papers, including an annotated bibliography (Dennis, 1984), describing the use of micro-computers in field research. However, as the editors noted, the basic tasks for which computers were being used were "storing, retrieving, sorting, cross-indexing, and in some cases, counting." (Conrad and Reinharz, 1984:8). Computers were seldom being used to statistically analyze textual data. More recent contributions to this methodological discussion report greater speed, scope and flexibility, but essentially the same research functions being performed by computers (Gillespie, 1986; Shelly and Sibert, 1986; Richards and Richards, 1987). Computerization has not led to widespread quantification.

There is, of course, no reason why the adoption of computer technology should lead to the displacement of qualitative by quantitative methods. But there are times when the latter could usefully complement the former. However, many field researchers, computer advocates among them, have a deep suspicion of quantitative methods.

> *"Researchers who specialize in nonnumeric data should begin to harness the text processing and interpreting power of the computer. This does not mean that they should sacrifice the information content of their data to inappropriate numerical manipulation. The computer has, unfortunately, already contributed to the overnumericization of much social science research. People engage in inappropriate transformation and analysis of nominal data because packages are available. People inappropriately structure their data into filled rectangular matrices because that is the structure demanded by analysis programs. People even ask their respondents questions in a format (fixed choice, single response) amenable to computer processing rather than to human processing."*
> *(Sproul and Sproul, 1982:288).*

As the following discussion reveals, we admit, but with few apologies, to all of these "inappropriate" research strategies. Our position, hardly a radical one, is that methods should be chosen on the basis of one's research question and the nature of one's subjects, rather than on *a priori* grounds. We chose semi-structured interviews as our primary data collection method,

with short fixed-choice questionnaires as a secondary method, because we believed interviews were the most appropriate method of data collection for this particular study. We also felt that quantification of the interview transcripts would assist our analysis and understanding of the qualitative data. Nevertheless, we do recognize that quantification of textual data can, in some cases, be inappropriate and counter-productive. Hence, the warnings above are taken seriously in this description and evaluation of a method-merging experiment.

RESEARCH DESIGN

Our data come from a 1985 study of high school dropouts in Edmonton, Alberta. The study was designed to provide insights into reasons for leaving school, relationships with family and friends following school exit, job search and employment histories, and the experience and consequences of unemployment. Data were collected via tape-recorded semi-structured interviews with dropouts who were also asked to complete a short, self-administered questionnaire. These two methods were supplemented with a limited number of unstructured interviews with key informants such as high school teachers and guidance counselors, and social service personnel involved in life skills training and/or career counseling.

Given the desire to generalize beyond our subjects, a random sample would have been preferred. However, such a research design was impossible to implement since a complete and up-to-date sampling frame of early school leavers could not be obtained or constructed. Instead, a non-random purposive sample of 168 young people who had dropped out of high school sometime in the previous few years was constructed. Referrals were obtained from schools, social service agencies and a corrections centre, employers and career counselors, and other sample members. An attempt was made to balance female and male, and employed and unemployed subjects in the sample. A total of 162 usable interview transcripts and completed questionnaires were obtained in this manner.

Following a referral, a potential respondent was contacted and provided with a one-page description of the study and an invitation to participate. If willing, the subject would be

interviewed (at the time, or at a later date) in a public place (usually libraries, restaurants, or shopping malls) by a member of the research team or by one of half a dozen trained interviewers. Virtually all contacted individuals agreed to participate in the study. The tape-recorded interviews ranged in length from around 30 minutes to over an hour.

Semi-structured interviews were chosen as the primary means of data collection because, it was felt, they could best illuminate the experiences and opinions of these young people. While there were clearly some central issues which the research team wished to address, semi-structured interviews could incorporate new questions as the study proceeded. An interview guide containing lead questions and "probes" was used to ensure that nine key topics were covered. While interviewers were instructed to address each of these areas, preferably in the order they appeared in the guide, they were free to add questions, move away from sensitive areas, reorder and rephrase questions, and follow up new leads.

Following the interview, respondents were asked to complete a short questionnaire. This research instrument included basic demographic questions. It also provided a cross-check on information meant to be obtained through the interviews such as the highest grade completed in school and current employment status. In addition, it allowed the use of a few standardized attitudinal questions about self-esteem, the value of education, and employment goals. Since this study of dropouts in the labour market was part of a larger research project including surveys of high school and university graduates, a limited number of comparisons across these samples were desirable. Finally, one of the research questions asked about the relationship between unemployment and crime. Previous research suggested that a self-reported, anonymous check-list of deviant acts was the best approach for collecting such information.

Although some of the questions in the interview and in the questionnaire addressed potentially sensitive subjects (school and work problems, relationships with parents, deviant behaviour), respondents appeared comfortable answering them, or skipping those they did not wish to answer. To ensure anonymity, all identifying information about respondents was destroyed after interviews were completed. The taped interviews were subsequently

transcribed directly into individual micro-computer files. These files were identified by an interview number linking the transcript to the questionnaire completed by the same respondent. Printed copies of each transcript were also produced.

QUANTIFYING INTERVIEW DATA

The printed interview transcripts have been used extensively (Krahn and Tanner, 1989; Samuelson, 1989; Tanner, 1990), and further analysis continues. Direct analysis of the computer files containing the transcripts has been undertaken less often, perhaps because of lingering attachments to pen and paper on the part of some research team members. Computer analyses of the textual data have not been particularly sophisticated, involving little more than marking of key passages for subsequent scanning and retrieval, embedding of some interpretive comments in the text, and efficient copying of previously marked quotes.

However, the interview data have also been statistically analyzed (Hartnagel and Krahn, 1989; Samuelson, 1989), and it is this process which we evaluate below. The desire to link interview and questionnaire data from the same subjects, and the standard sequence of questions in the interviews, suggested the possibility of transforming the transcript material into a rectangular numeric data set with rows representing respondents and columns containing variables. Such a research strategy is bound to be criticized for, indeed, it involves reducing "rich" interview data to numbers for the sake of statistical analysis. In defence, we reply that the numeric data set was constructed to take better advantage of the questionnaire data, but only as a supplement to the textual material. The exercise was also treated as a methodological experiment, requiring cautious use and careful evaluation.

Two members of the research team and two of the additional interviewers each read a large number of the transcripts, attempting to identify the types of answers being provided. This proved to be a very difficult task, as anyone familiar with interview data could predict. The semi-structured interview format meant that information relevant to a particular question (e.g. *how did your family react to you quitting school?*) might appear where it would be expected in

the transcript. It could also arise elsewhere as the conversation shifted from topic to topic. The answer might be clearly stated, or it might be hidden within a long and rambling discussion of only partially-related topics. It might not be found at all if the interviewer had failed to ask this particular question or if the respondent had declined to answer it.

This initial reading of transcripts was followed by hours of equally difficult, but highly useful, discussion regarding coding strategies. Ultimately, two basic sets of variables - sorting and analytic - were defined. Sorting variables, binary in form, were meant only to identify interview transcripts containing textual material relevant to a particular theme or research question. For example, such variables could "mark" transcripts containing information on why jobs had been terminated, opinions about whether *society owes you a job*, or respondents' volunteered discussions of criminal behaviour.

These sorting variables were not designed for statistical use. Instead, anticipating subsequent interest in certain types of text but encountering problems in developing a coding scheme, they were devised to serve the simple function of quickly locating relevant transcripts. In practice, the SPSSx program used to process the data set could be asked for the ID numbers of all interviews in which a specific topic was discussed. The researcher could then read these transcripts and, if she or he wished, leave thematic markers at key points. If the transcripts were read via micro-computer, some of the ethnographic software applications noted earlier could be used to imbed electronic tabs within the text.

The several different types of analytic variables were, however, clearly intended for statistical analysis. The simplest were also binary measures, representing affirmative or negative answers to questions like: *have you ever been unemployed (before)?*, *are you currently looking for a job?*; and *are you supporting supporting anyone else?*. Nomimal variables, with several or more possible values, were constructed from answers to questions like: *where are you currently living?* and *who do you (usually) spend your free time with?* Depending on the question, multiple answers might be common, requiring the specification of up to three variables for the first, second and third answers given. For example, *how do you look/have you*

looked for work? elicited a fairly large range of multiple answers, requiring the construction of three variables and a coding system with thirteen different categories (newspaper; door-to-door; family contacts; telephone; through volunteer work; and so on).

Answers to some of the questions in the interview guide allowed the ranking of respondents on various dimensions. Ordinal variables were constructed from responses to questions like: *did any of your friends quit school?* (none; some; many); *would you go back to school?* (no; maybe/unsure; yes); and *how have you felt lately physically?* (not very good; OK/getting by; quite/very good). While only a few subjects answered in exactly these words, such categories allowed relatively easy ranking of most respondents.

However, some ordinal variables required coders to interpret the text or read between the lines. Examples included *how satisfied does the respondent appear with his or her job?* (dissatisfied; hard to tell; satisfied) and, after asking the respondent to list all previous jobs, *how long did the respondent's jobs usually last?* (generally short-term; some short-term and some longer; generally long-term). Finally, an attempt was made to create variables at the interval level of measurement from responses to questions about countable events. *How long ago did you leave school?* and *how long have you been unemployed/at your job?* are examples.

A final pre-coding task involved the designation of "missing data" codes. This is a necessary step in all survey research but was even more critical in this project. While the goal was to present each respondent with the same or similar questions, the use of semi-structured interviews meant that some questions would be omitted in almost every interview. Hence, in addition to a standard code for non-response, a second would be needed to note where a question had not been asked. While in some analyses the distinction between these two types of missing data would be of little significance, in others it could be important. We also allowed coders to use a third "irrelevant answer" code for situations where the question had been asked and an answer was provided, but where the respondent discussed something else.

Once the variables were defined and coding rules were established, the transcripts were systematically coded in what amounted to a content analysis of the interview text. Even with carefully specified coding rules, it is often difficult to code textual data without considerable, knowledge of the social context from which these data originate (Carley, 1988). This problem was reduced by having two members of the research team actively involved in the coding of the interview transcripts. Three of the four other coders had also participated in the interviewing, but these individuals did not code their own interviews.

Ten of the transcripts were coded a second time, by a different coder, to provide a small test of inter-coder reliability. When the coding of the textual data was complete, the results were merged with the data previously retrieved from the self-administered questionnaires. An SPPSx program was then written to process and analyze this numeric data set containing 213 variables, about 75% of which had been constructed from the textual data obtained in the semi-structured interviews.

ASSESSMENT

Data collection procedures can be assessed in terms of "information adequacy", or the quality of the data, and "efficiency" or cost (Zelditch, 1962:569). Added to this should be "research utility", the degree to which the data, once collected, are of value in answering research questions. The same three criteria can be used to assess the data transformation procedure described above. We begin with information adequacy, since conclusions about research utility and cost efficiency are obviously contingent on the quality of the data.

A. INFORMATION ADEQUACY

1. Missing Data

When analyzing textual data, the absence of information on a particular subject in any one transcript is regrettable. If, however, this topic was discussed in a reasonable number of other interviews, the subject can still be explored. But once textual data have been quantified,

the norms of survey research begin to take precedence and even small amounts of missing data may be problematic. Unfortunately, the potential for missing data in a numeric data set built from semi-structured interviews is considerable. Not all of the questions in the interview guide would have been asked of each respondent. Furthermore, if a respondent did not understand the question, it might have been altered. If someone appeared reluctant to answer, there would have been little further effort made to solicit an answer.

We found missing data problems resulting much more often from questions not being asked than from respondents' unwillingness to answer. For example, only 67% of the employed respondents discussed what they disliked about their job. In all but four of the cases where this information was absent, the question had not been asked, for a variety of conversational reasons. However, 97% of the sample provided an answer to *why did you leave school?*, a question asked of everyone. Considering all of the variables created from textual data, about one-third had a 90% or higher response rate. Another 41% had rates of response in excess of 80%, while 13% contained answers from between 70% and 79% of the sample. The remaining 13% of the variables had lower response rates.

Thus, in terms of semi-structured interviews, this was a relatively complete data set. The use of an interview guide and the prior training of interviewers helped ensure that most of the key questions were covered. But by survey standards, there were large gaps in the data set. Hence, when these variables are used in multi-variate statistical procedures, cumulative missing data problems quickly arise. For example, in one analysis the inclusion of five of these formerly textual variables in a multiple regression equation reduced the sample from 162 to 106 cases (Hartnagel and Krahn, 1989).

One solution to such problems would be to build even greater structure into these interviews or, in order words, to move even further toward standardized survey research methods. However, we deliberately chose semi-structured interviews and, if anything, would have preferred if some of our interviewers had pursued more of the interesting conversational leads instead of sticking closely to the interview guide. Thus, for our purposes, further

structuring of the interviews would not have been a solution.

Instead, we have responded to the missing data problem by omitting some of the more problematic variables in multi-variate analyses or by constructing indices out of several partially complete variables in order to cover more cases. Equally important, we try to examine our results to determine whether they might be biased due to a significant reduction in the sample. Most important, we attempt to remind ourselves that this study was not designed as a survey. We should perhaps program our computer to print a caution at the top of each page of output: *These numbers were once conversations.* Our advice to researchers uncomfortable with these solutions is not to attempt the quantification of textual data from semi-structured interviews.

2. Inter-Coder Reliability

Inter-coder reliability was simply operationalized as the total number of agreements between coders divided by the total possible number of agreements. Ten of the transcripts were double-coded and measures of reliability were then calculated for each variable. Although ten pairs of values are a small sample with which to assess reliability, these limited tests can still point to some of the coding problems which emerged.

Miles and Huberman (1984:63) suggest that inter-coder reliability should reach 70% at the outset. As coders gain experience, 80% or 90% should be the goal. We found 75% agreement across the full set of variables created from textual data. We had expected some types of variables to exhibit higher reliability than others, and did find differences across types. However, there was considerably more variation within these groups. The lowest reliability for a single variable was 40%, but a reasonable number exhibited 100% reliability. In other words, they were coded similarly by two coders in all ten sets of double-coded transcripts. Thus, excluding the more problematic variables, our limited test suggests a reasonable level of inter-coder reliability.

a. Sorting Variables

Sorting variables have a simple function - to identify transcripts containing discussion of specific topics. Hence, we expected them to exhibit considerable inter-coder reliability. Some functioned very well (e.g. *Is information provided about wages?*). But, on average, the sorting variables showed somewhat lower than average reliability. For example, respondents were asked: *Did school help you look for work?* An analytic binary variable was used to code responses to this question, and then a sorting variable was constructed to identify interviews where reasons for either a postive or negative answer were elaborated. The following exchange is from one of the three transcripts where coders disagreed about whether such reasons were provided.

[Interview 30; male]

I: Okay, to get back to your school then, do you think that it helped you find the jobs that you have found?

R: What do you mean?

I: Like finding the job at MacLeod's, do you think school helped you find that job, your education that you received at school?

R: Not really.

I: No?

R: Didn't really need no education for that, it was just looking up numbers in a book and finding them on the shelf and putting them there.

This young man claimed that school had not helped him look for work. One of the coders concluded that he provided a reason for his negative answer -- low-level jobs available to dropouts don't require much education. The other coder may have reasoned that the respondent was talking about the nature of the job, not about his job search experiences. Personally, I would have "marked" this transcript. The exchange is relevant to debates about the value of educational credentials in youth labour markets, but one of the coders did not recognize the theoretical significance.

The problem is that even though coders were not classifying the content of answers, they still had to interpret text to decide whether it was relevant to the issue in question. Different coders might have very different interpretations. Inter-coder reliability was also low on several sorting variables which asked: *is further useful information provided about ....?* Again, coders tended to have different opinions about whether a passage was really useful.

The solution might lie in even simpler sorting variables such as *does the respondent say anything else about ...?* But even here coders might disagree, since conclusions about the relevance of tangential comments must still be theoretically informed. In fact, intra-coder (over-time) reliability could also be low for such variables. On re-reading a passage of text, one's assessment might differ because of new insights obtained from other sources.

In short, sorting variables have an important non-statistical function in data sets such as this. They point to transcripts in which specified topics are discussed. To ensure that they function effectively, they should be kept as simple as possible. When additional coders are involved, researchers should take the time to explain the issues underlying the questions and the key themes meant to be highlighted in the text.

b. Analytic Variables - Binary

Analytic binary variables revealed a better than average level of reliability, with a considerable number being coded identically by the pairs of coders in all ten test cases. These variables were most reliable when they simply reflected the presence or absence of some respondent characteristic (e.g. *Have you ever been unemployed?*). But where some interpretation of text by the coders was required, reliability began to suffer. The interview guide contained the following question in a section about financial resources: *Were you getting any (financial) help in the past?* The binary variable constructed from respondents' answers to this question exhibited only 50% inter-coder reliability. One of the conversations over which coders disagreed went as follows:

[Interview 89; male]

I: What do you live on?

R: The money from my jobs.

I: How about some financial help from your parents?

R: I borrowed sixty bucks about three months ago off of my Mom and Dad but that's it.

This coding disagreement probably centres around whether a "loan" from parents is financial help. I would consider it to be, but failed to communicate this interpretation to one of the coders. Again, as with sorting variables, the interpretive rules must be clear and consistent. This obviously applies to situations where several or more coders are translating text into numeric code. But it is also applicable to the solitary researcher. The coding of interview transcripts is a time-consuming task. Without a clear set of coding rules, intra-coder reliability could also suffer as a researcher worked through a large set of transcripts.

c. Analytic Variables - Nominal

Nominal analytic variables, intended to categorize different types of answers, were somewhat more problematic, unless the question was fairly simple (e.g. *Who are you living with?*). But quite often the questions we put to our respondents were more complex. Coders had to think carefully about the categories in which they might place an answer. We were interested, for example, in where dropouts attributed the blame for high rates of joblessness among youth. After discussing unemployment for several minutes, we asked: *Whose responsibility is youth unemployment?* A set of seven codes was developed, including government, business, youth, and so on. We also allowed an "irrelevant answer" code for situations where the subject was not responding to the question. Inter-coder reliability for this particular variable was only 60%.

[Interview 22; female]

I: ..... Whose responsibility do you think it is then?

R: Unemployment? You know I do get a little ticked off at the government for creating make-work programs that they pay someone $3.80 an hour when nobody could afford that anyway, and they can

bring in their surveys and say 'look what we've done'. I really, I think there's a lot of bullshit going on. Let them take the $3.80 an hour jobs. I mean, you have your set guidelines, government's STEP in place, make $5.50 starting, anybody else makes $3.80 starting. Something is screwed up somewhere there, why not give them all $4.50 and call it even.

One of the coders decided this individual was blaming the government while the other felt the answer was not addressing the question. I agree with the latter, since this individual was criticizing the government's implementation of youth job-creation programs, [2] not blaming the government for the problem to which it was responding.

Perhaps we should not have tried to code answers to such abstract questions. However, some respondents wasted little time in identifying a culprit, and their answers were easily coded. For others, like the young woman quoted above, answers that wandered off to other issues suggested that the attribution of blame for youth unemployment was not part of their everyday thinking. This is an important observation if one wishes to evaluate the utility of attribution theory for explaining the opinions and behaviour of youth. Once again, with clearer coding rules, inter-coder reliability for such variables might have been improved.

In a number of situations, we developed sets of two or three nominal variables to code multiple answers. When asked *Did you have any problems when looking for work?*, most respondents identified more than one. Hence, we created three variables (first, second and third problem mentioned) and a coding scheme with three basic types of problems (personal feelings; labour market; lack of resources) sub-divided into 19 distinct categories. Inter-coder reliability, calculated across all three variables was high (92%), no doubt because the opportunity to code more than one answer reduced inter-coder discrepancies created by being forced to choose among several possible coding categories. [3]

When we failed to create enough variables to accomodate the answers being given, reliability declined. Interviewees were asked *How do you spend your time when you are not working (looking for work?)* We allowed only two variables, with 17 coding categories, for answers, and inter-coder reliability was low (55%). For example, one of the coders recorded

reading and movies from the following exchange, while the other coded reading and visiting with friends.

[Interview 20; male]

I: Okay, how do you spend your time when you're not at school?

R: Ah, it ranges - anything from reading to staring at the TV to cross-country skiing, and evenings are just movies or concerts.

I: ..... Who do you spend your time with usually?

R: Um, usually with my girlfriend, and sometimes with the crowd, all the friends; sometimes, I just sit at home.

I: Where do you go with your other friends - same kinds of things?

R: Um, with the other crowd of friends, it's mainly the same group I used to chum with when I was in High School, so that's the bar crowd.

I: So you usually go to theatres with your girlfriend?

R: Yeah, it's usually just us.

Construction of a third variable would have allowed the coding of more activities and presumably would have led to higher inter-coder reliability. However, coding problems would have remained, since this respondent listed more than three activities. An alternative would have involved creation of broader categories such as group versus individual activities. The drawback of this strategy is that data users would not be able to look at the detailed codes and create their own broader categories. Our recommendation would be to use detailed codes and to create a reasonable number of variables for multiple answers, recognizing that a small minority of respondents will always exhaust the coding possibilities provided. In such cases, the theoretical interests of the principal researcher(s) would dictate which answers were not coded.

d. Analytic Variables - Ordinal

These ranking variables showed average but widely varying reliability. Once again, clearer instructions would have reduced some of the coder disagreements. However, inspection of the

difficult to code transcripts suggests that some of these variables might have been better left as binary measures. For example, three categories (not very well; OK; very well) were provided to classify answers to *How do you get along with your family?* Inter-coder reliability was low (50%). In four of the five discrepant cases, including the one below, the first coder chose "OK" while the second picked "very well".

[Interview 20; male]

I: Okay, maybe we'll talk a little bit about how you get along with your family?

R: Pretty good, actually. I think my Dad is getting a little bit....he's kinda wondering about what I'm planning on doing, if I'm gonna be staying home until I'm 40. So he's getting a little bit uptight about that which gets Mom a little bit uptight 'cause Dad's uptight, but it's no real problem. Everytime he gets like that, it just seems like coincidence, I'm drumming up something like this Athabasca University thing so he's contented again.

I: So, how about your Mom?

R: Ah, as far as my business goes, she usually stays out of it, out of my personal life whether it be school, or my girlfriend. She doesn't get involved that much unless I mention something to her. I wouldn't call it a problem at home.

The two coders were probably working with different personal views of what constitutes a harmonious family. If I had been coding this exchange, I would also have been unsure. Given my own ambivalence, a binary variable (not very well; OK or better) would have been more useful. Fortunately, data analysis programs allow easy recoding of variables, and this transformation can still be done while analyzing the data. But not all of the ordinal variables exhibited low inter-coder reliability. Answers to the question *Would you like to get more education?* (yes; maybe; no) were coded identically for each of the ten double-coded transcripts. Thus, our advice would not be to avoid ordinal variables when coding textual data, but to avoid building more variation into the coding scheme than the text allows.

And how does one know, without the benefit of hindsight and tests of inter-coder reliability, when a higher level of measurement is appropriate? There is no easy answer, but Bryman (1988:145) provides some useful advice: do not go beyond categories that

respondents themselves would use. One can imagine young people answering in positive, ambivalent, or negative terms when asked about future education plans. Hence, an ordinal measure might be useful for coding their answers. But negative and qualified positive answers may more or less exhaust the range of replies to a question about getting along with parents. Thus, in this case, trying to construct an ordinal measure out of essentially binary responses may be inappropriate.

e. Analytic Variables - Interval

The relatively few interval-level variables we constructed showed average reliability. Our advice, again, is: a) do not build more variation into the numeric data than its textual source allows; and b) work with categories the respondents themselves might use. An interval variable intended to code answers to *How many of your friends have been unemployed?* exhibited 90% reliability. Young people can and will count when asked such a question. However, they are considerably less precise when asked *How long ago did you leave school?* We instructed the coders to record the number of months since the interviewee had left school, and found only 40% inter-coder reliability in the ten test cases. Examination of several of the "problem" transcripts is useful.

[Interview 19; female]

I: So, how long ago was it that you quit school? Three or four years ago?

R: I'm 20 now, I was 15-16, ...

[Interview 35; female]

I: So you dropped out in Grade 10, when was that?

R: Oh, about five, six years ago.

[Interview 59; male]

I: Okay, how long ago did you finish then?

R: Just last year.

I: So you've been out a year?

R: Yeah, almost.

The level of precision we had hoped to obtain simply did not exist. To be fair to ourselves, we tried to code the data in months because we wanted to be able to differentiate between those who were only a few months out of school and those who left a longer time ago. In retrospect, an ordinal level of measurement with a reasonable number of categories would have been more appropriate and equally useful. As it was, all of the coding discrepancies were small (e.g. 60 versus 66 months) and we can still construct ordinal variables from our overly-ambitious interval measures.

B. RESEARCH UTILITY

Transcription of the interviews directly into micro-computer files was clearly a useful research strategy, and not only because the typing was completed more quickly. The resulting files could be printed on paper but were also available for analysis with software designed for textual data. In addition, they were highly portable. But was there research value in quantifying the textual data?

In some analyses, we have used the SPSSx computer program simply to locate interviews in which particular topics were discussed. Having employed the sorting variables in the way they were intended to be used, we then read the relevant transcripts and extracted useful quotes. For example, a statistical analysis of the effects of unemployment on crime was supplemented with a discussion of the dropouts' own views on the subject (Hartnagel and Krahn, 1989). A sorting variable identified those transcripts containing relevant discussions. The earlier examples used to highlight sources of low inter-coder reliability were identified in much the same way -- the computer was instructed to find cases where the two coders had disagreed about how to code a specific variable.

While our system works well enough for such purposes, it is really rather crude and inefficient. If sorting of transcripts had been our primary analysis goal, there are a number of

much more flexible micro-computer applications which could be used for the same purpose. [4] However, we had also hoped to undertake statistical analyses with the transformed data set once it had been linked with the questionnaire data. It is here that the value of quantifying the interviews is more apparent.

The ability to quickly report that 51% of these dropouts had been skipping classes on a regular basis prior to quitting school, or that 75% came from families where someone else had been unemployed in the past (Krahn et al., 1985), has been very useful for summarizing findings from this study. Frequency distributions are also valuable for demonstrating that impressions obtained from qualitative analyses are, in fact, soundly based. In our reading of the transcripts we were impressed by how few of these dropouts seemed to be truely alienated from the education system. Our SPSSx program quickly confirmed this impression, revealing that a large majority (?0%) answered positively when asked if they would like to get more education (Krahn and Tanner, 1989).

We have also found that qualitative analyses can be effectively supplemented by cross-tabulations based on the numeric data. For example, the impression that females and males were giving fairly similar reasons for leaving school could be easily confirmed by cross-tabulating reasons for leaving by gender (Tanner, 1990). The hypothesis that older dropouts would evaluate their (low skill, low pay) jobs less positively than would younger dropouts could be quickly tested (Krahn, 1991). [5]

The creation of a standardized numeric data set has proved to be equally valuable in situations where we wished to undertake more complex statistical analyses. For example, in an examination of the effects of unemployment on crime, five of the variables created from interview text were used as independent variables in a series of multiple regression analyses (Hartnagel and Krahn, 1989). The dependent variables were indicators of alcohol and drug use, and of criminal behaviour, from the self-report questionnaire. Given the theoretical developments in this research area, analyses of the unemployment-crime relationship that omitted key control variables and that did not use multiple indicators of labour market problems,

would be of little value. [6]

Critics might note that a similar analysis could have been done without the use of qualitative data. *"If you want to use survey analysis techniques, why bother with semi-structured interviews?"* We bothered with them because we felt they could best provide the information we wanted. We then went on to quantify the text to allow analysis strategies that would otherwise be very difficult or impossible to implement. In this example, the interview material supplemented the quantitative analysis. In a second assessment of the same research question, the textual data were used much more extensively (Samuelson, 1989). Ten of the least and another ten of the most deviant sample members were randomly selected. The interviews completed with these individuals were then carefully examined to see whether clear patterns of explanatory variables which would discriminate between these groups could be observed.

We conclude that the quantification of these interviews had considerable research utility. If sorting of transcripts had been our only goal, other computer techniques would have been more useful. But the approach described above becomes more valuable as the need for more complex statistical analyses, particularly with control variables, increases. The practicality of this approach is, in large part, a function of the degree to which the interviews are structured and standardized. It would be much more difficult to code unstructured interviews into a standardized numerical data set. Finally, the research value of this approach increases with sample size. This is the case for most research methods, but it is particularly important here since the greater amount of missing data quickly reduces sample sizes in multi-variate analyses.

C. EFFICIENCY

Excluding time spent by the research team in designing and managing the study, interviews consumed about three hours each. The taped discussions themselves averaged under an hour in length. The rest of the time was needed for contacting potential respondents, travelling to and from the chosen meeting place, repeating this exercise when respondents failed to appear, and reporting back to the interviewing supervisor. Transcription of the

interviews into micro-computer files (by skilled transcribers) also averaged around three hours per interview.

The subsequent coding of these transcripts took somewhat less time, slightly more than two hours per interview. This excluded the time spent on a number of related tasks: developing the coding scheme and creating a codebook; recoding ten of the transcripts for the test of inter-coder reliability; keypunching and cleaning the data; and writing the SPSSx program. While these additional tasks were not timed, we estimate that, including coding, the transformation of the interviews into a numeric data set added about 50%, that is, another three hours per interview, to the time spent on interviewing and transcription. If one was paying for this work, as we were, research costs would be expected to go up by an equivalent amount.

Was this, then, an efficient research procedure? We think it was. The data appear to be of reasonable quality, or better. We have been more than satisfied with the research value of both the textual and numeric data sets, despite some of the coding and analysis problems encountered. Obviously, a leaner research budget might have kept us from coding the transcripts. But we would not have reconsidered the use of micro-computers for transcription itself. Given its greater speed compared to traditional typing, computer use at this stage was itself a cost-saver. The portability of data files, and the many options for computer-based analysis of textual data, are bonuses.

SUMMARY AND CONCLUSIONS

In an excellent ethnography describing the shaping of aspirations of young men growing up in a poor neighbourhood, MacLeod (1987:165-85) provides a detailed and helpful account of his research methods. After wrestling with his field notes and interview transcripts for some time, he found that his most useful data organizing tool was a large chart on which he listed information on each of his subjects: race, peer group, school performance, expressed aspirations, parent(s)' education and employment, alcohol and drug use, and so on. "Once the data had been systematically laid out, patterns began to emerge." (MacLeod, 1987:185)

If this researcher had taken the small further step of using numbers rather than words to record his information, he would have produced a rectangular data set much like the one described above. And if he had studied more than 15 young men, he might well have been tempted to computerize his data chart. In other words, when qualitative researchers begin to develop causal explanations of social phenomena they often employ the same logic of analysis used by quantitative researchers. They examine *patterns* in their data; survey researchers look at *relationships*. Thus, the quantification of textual data from semi-structured interviews simply provides qualitative researchers with an additional research tool.

If qualitative researchers approach quantification carefully and cautiously, the construction of a numeric data set can be an efficient and useful research strategy. Our experience with the procedure highlights the different research functions of sorting and analytic variables. The former should be kept as simple and non-interpretive as possible. Coding rules for both types of variables should be clearly specified, particularly if several coders are involved in transforming the text into numbers.

For nominal variables where multiple answers might be found, detailed codes and a reasonable number of variables should be created. While ordinal and interval level variables are very useful, researchers should avoid building more variation into the numeric data set than its textual source allows. They should also try to work with categories the respondents themselves might use.

Having quantified their interview data, researchers should be prepared to deal with substantial missing data problems. Equally important, they should regularly remind themselves that the numbers they are analyzing originated in semi-structured conversations. Quantification of textual data is a useful supplement to traditional methods of analysis, but not a replacement for them.

NOTES:

1. These response rates exclude variables containing second or third answers to a question generating multiple responses. For example, 88% of the sample provided an answer to *what kind of advice would you give someone looking for a job?*, but only 52% went on to give a second piece of advice. Only 4% of the variables had response rates of less than 50%, with the lowest being 35%. This was a question about total income per week. It was not asked of most of the unemployed respondents (almost one-half of the sample) since it generally appeared inappropriate, given that the money they were receiving was not really "income". Other questions about *what do you live on?* had much more complete data.

2. STEP is the acronym for Student Temporary Employment Program, a program which subsidizes the hiring of young workers. Government departments were hiring young people through this program, but frequently "topped up" the hourly rate of pay to $5.50. This respondent appears to be criticizing the pay discrepancy between these particular STEP jobs and others in the private sector.

3. If only two answers were provided, a non-response code was assigned to the third variable. Inter-coder reliability was calculated across the three variables. If one coder placed a particular answer into the first variable while the other used the second variable for this answer, the coding was still considered consistent.

4. These applications would, of course, still require prior coding. Pertinent passages would be tagged with electronic markers in the text. With our less flexible approach, interviews containing useful information would be identified with a sorting variable in a second numeric data file, but the exact location of the relevant passages would not be recorded.

5. It should be noted that frequency distributions and simple cross-tabulations can also be generated by some of the micro-computer software packages developed for ethnographic data (e.g. Richards and Richards, 1987).

6. The variables constructed from interview data were: total number of months since leaving school; total number of months of unemployment; total number of jobs; personal financial status; and living arrangements. Two other key independent variables, gender and current employment status, were recorded in both the questionnaires and the interviews.

REFERENCES

Bryman, Alan
1988 Quantity and Quality in Social Research. London: Unwin Hyman.

Carley, Kathleen
1988 "Formalizing the social expert's knowledge." Sociological Methods and Research 17(2):165-232.

Connidis, Ingrid
1983 "Integrating qualitative and quantitative methods in survey research on aging: an assessment." Qualitative Sociology 6(4):334-52.

Conrad, Peter and Shulamit Reinharz
1984 "Computers and qualitative data: editors' introductory essay." Qualitative Sociology 7(1 and 2):3-15.

Dennis, Deborah L.
1984 "'Word crunching': an annotated bibliography on computers and qualitative data analysis." Qualitative Sociology 7(1 and 2):148-56.

Gillespie, Gilbert W.
1986 "Using word processor macros for computer-assisted qualitative analysis." Qualitative Sociology 9(3):283-92.

Hartnagel, Timothy F. and Harvey Krahn
1989 "High school dropouts, labour market success, and criminal behaviour." Youth & Society 20:416-44.

Krahn, Harvey
1991 "Coming to terms with marginal work: job evaluations of young high school dropouts." Unpublished paper, Department of Sociology, University of Alberta.

Krahn, Harvey and Julian Tanner
1989 "Bringing them back to school." Policy Options Politique (March):23-24.

Krahn, Harvey, T.F. Hartnagel, G.S. Lowe, J. Tanner and L. Walter
1985 Youth Employment, Underemployment and Unemployment: A Study of the Transition from School to Work in Edmonton. Population Research Laboratory, Dept. of Sociology, University of Alberta.

MacLeod, Jay
1987 Ain't No Makin' It: Leveled Aspirations in a Low-Income Neighborhood. London: Tavistock.

Miles, Matthew B. and A. Michael Huberman
1984 Qualitative Data Analysis: A Sourcebook of New Methods. Beverly Hills: Sage.

Richards, Lyn and Tom Richards
1987 "Qualitative data analysis: can computers do it?" Australian and New Zealand Journal of Sociology 23(1):23-35.

Samuelson, Leslie
1988 The Out-of-School Experiences of Dropouts: Labour Market Success and Criminal Behaviour. Unpublished PH.D. thesis, Department of Sociology, University of Alberta.

Shelly, Anne and Ernest Sibert
1986 "Using logic programming to facilitate qualitative data analysis." Qualitative Sociology 9(2):145-61.

Sieber, Sam D.
1973 "The integration of fieldwork and survey methods." American Journal of Sociology 78(6):1335-59.

Sproull, Lee S. and Robert F. Sproull
1982 "Managing and analyzing behavioral records: explorations in nonnumeric data analysis." Human Organization 41(4):283-90.

Tanner, Julian
1990 "Reluctant rebels: a case study of Edmonton high school dropouts." Canadian Review of Sociology and Anthropology 27(1):74-94.

Wallace, Claire
1987 For Richer, For Poorer: Growing Up In and Out of Work. London: Tavistock.

Zelditch, Morris Jr.
1962 "Some methodological problems of field studies." American Journal of Sociology 67(5):566-76.

RESEARCH DISCUSSION PAPER SERIES

58. The Effects of Religious Factors on Childlessness: The Canadian Case.
Vijaya Krishnan. April 1989.

59. A Study of the Contemporary Population Migration in the People's Republic of China.
Yihua Yuan. May 1989

60. Might Suicide Underreporting Affect Hypothesis Test? The Case of Suicide and Divorce.
Mark Speechley and Frank Trovato. July 1989.

61. Demographic Concerns in Belles-Lettres: Insights From the Writings of Günter Grass and Margaret Atwood.
Karol Krótki and Guy Germain. January 1990.

62. Some Lacunae in Demographic Data Collection in South Asia with Special Reference to India.
Parameswara Krishnan. January 1990.

63. The Emergence of Multiethnicities in 1981 and 1986: Their Sociocultural Significance.
Karol Krótki and Dave Odynak. February 1990.

64. Dynamics of Internal Migration in Bangladesh.
A.K.M. Nurun Nabi. April 1990.

65. Why Are Canadians Dying Out?
Karol Krótki. April 1990.

66. A Theoretical Model for the Study of Migration and Urbanisation
David Achanfuo-Yeboah. May 1990.

67. Political Demography of the Emergence of Bangladesh
A.K.M. Nurun Nabi and P. Krishnan. May 1990.

68. Dependency Ratios: Conventional and Real.
Parameswara Krishnan and Manoshi Das. June 1990.

69. Log-Linear Modelling and Simultaneous Standardization of Rates.
Frank Trovato. November 1990.

70. Unemployment and Occupational Aspirations: A Panel Study of High School Graduates.
Susan Empson-Warner and Harvey Krahn. November 1990.

71. An Analysis of Native Mortality in Canada.
Andrea Werner-Leonard and Frank Trovato. November 1990.

72. Ethnic Variations in Fertility: Microeconomic and Minority Group Status Effects.
Tina W.L Chui and Frank Trovato. December 1990.